1423

NPS-53-89-010

NAVAL POSTGRADUATE SCHOOL

Monterey, California

HOW GOOD ARE GLOBAL NEWTON METHODS?

Part 1

Allen Goldstein

February 1989

Approved for public release; distribution unlimited

Prepared for:

Naval Surface Weapons Center
Dahlgren, VA 22448-5000

NAVAL POSTGRADUATE SCHOOL

Department of Mathematics

Rear Admiral R. C. Austin
Superintendent

Harrison Shull
Provost

This report was prepared in conjunction with research conducted for the Naval Surface Weapons Center and funded by the Naval Postgraduate School.

REPORT DOCUMENTATION PAGE

REPORT SECURITY CLASSIFICATION: UNCLASSIFIED	1b RESTRICTIVE MARKINGS
SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION/AVAILABILITY OF REPORT: Approved for public release; distribution unlimited
DECLASSIFICATION/DOWNGRADING SCHEDULE	
PERFORMING ORGANIZATION REPORT NUMBER(S): NPS-53-89-010	5 MONITORING ORGANIZATION REPORT NUMBER(S): NPS-53-89-010
NAME OF PERFORMING ORGANIZATION: Naval Postgraduate School; 6b OFFICE SYMBOL (If applicable): 53	7a NAME OF MONITORING ORGANIZATION: Naval Surface Weapons Center
ADDRESS (City, State, and ZIP Code): Monterey, CA 93943	7b ADDRESS (City, State, and ZIP Code): Dahlgren, VA 22448-5000
NAME OF FUNDING/SPONSORING ORGANIZATION: Naval Postgraduate School; 8b OFFICE SYMBOL (If applicable): 53	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: O&MN Direct funding
ADDRESS (City, State, and ZIP Code): Monterey, CA 93943	10 SOURCE OF FUNDING NUMBERS: PROGRAM ELEMENT NO; PROJECT NO; TASK NO.; WORK UNIT ACCESSION NO

TITLE (Include Security Classification)

How Good Are Global Newton Methods. Part 1

PERSONAL AUTHOR(S)
Allen A. Goldstein

TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT (Year, Month, Day)	15 PAGE COUNT
	FROM 9/88 TO 1/89	1989 February 23	11

SUPPLEMENTARY NOTATION

COSATI CODES			18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Global Newton methods, unconstrained optimization, computational complexity

ABSTRACT (Continue on reverse if necessary and identify by block number)

ABSTRACT: 1) Relying on a theorem of Nemerovsky and Yuden(1979) a lower bound is given for the efficiency of global Newton methods over the class $C^1(\mu, \lambda)$ defined below.
2) The efficiency of Smale's global Newton method in a simple setting with a nonsingular, Lipschitz-continous Jacobian is considered. The efficiency is characterized by 2 parameters, the condition number Q and the smoothness S, defined below. The efficiency is sensitive to S, and insensitive to Q.

DISTRIBUTION/AVAILABILITY OF ABSTRACT: ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION: UNCLASSIFIED
NAME OF RESPONSIBLE INDIVIDUAL: llen A. Goldstein	22b TELEPHONE (Include Area Code): (408) 646-2335; 22c. OFFICE SYMBOL: 53Go

D FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

How good are global Newton methods. Part 1

A. A. Goldstein*

ABSTRACT: 1) Relying on a theorem of Nemerovsky and Yuden(1979) a lower bound is given for the efficiency of global Newton methods over the class $C^1(\mu, \lambda)$ defined below. 2) The efficiency of Smale's global Newton method in a simple setting with a non-singular, Lipschitz-continuous Jacobian is considered. The efficiency is characterized by 2 parameters, the condition number Q and the smoothness S, defined below. The efficiency is sensitive to S, and insensitive to Q.

KEYWORDS: Global Newton methods, unconstrained optimization, computational complexity

Global Newton methods are considered by some to be methods for minimizing a "strongly" convex function f defined on a real Hilbert space E. Strongly convex means that f is twice differentiable with a Hessian that is bounded from above and below. By $C(\mu, \lambda)$ we denote the set of all strongly convex functions whose Hessian is bounded below by μ and above by λ. The Hessian is invertible so that Newton's method is well defined for every point in E. Moreover a strong convex function achieves a minimum, where $\nabla f(x) = 0$. However Newton's method may not converge to a root of $\nabla f(x) = 0$ from arbitrary points in E. This is a *raison d'être* for the Global Newton methods. These methods, whose ingredients contain Newton steps, generate sequences that converge for every strongly convex functions and any starting point in E. The convergence rate is asymptotically superlinear. An early history of this subject may be found in Polak(1973), who cites contributions of Goldstein(1965), Pshenichnyi(1970), and Robinson(1972). More recent work is due to Bertsekas(1982), Dunn(1980), Hughes and Dunn(1984), and others. All of these results give estimated asymptotic rates of convergence. Global Newton methods for finding roots go back to at least 1934. They are related to continuation methods. An early history and discussion may be found in Ortega and Rheinboldt(1970,p235), who credit the basic idea to Lahaye(1934,1948). Current references may be found in Smale(1986-2). In general we regard a global Newton method as any algorithm incorporating Newton steps that that generates a finite sequence terminating in an approximate root. This is a point from which the ordinary Newton's method will converge. Other algorithms are available that terminate in an approximate root. The *efficiency* or iteration count of 2 such algorithms will be compared to a global Newton method. The word "algorithm" as used in this paper should be taken with "a grain of salt". We assume information that is not given with real

---

* supported by grants NIH RR01243-05 AND NPS LMC-M4E1

problems. Our excuse for doing this is that we hope thereby to gain insight and motivation for the future construction of good algorithms.

The efficiency of a Global Newton method was probably first analyzed by Kung(1976), using natural assumptions that imply a non-vanishing Jacobian. It appears that the next such result is due to Smale(1986-2) who established a global Newton method in the general setting of an analytic mapping between Banach spaces, both real or both complex. We revisit this problem below. Our assumptions are close to Kung's but our algorithm follows Smale. The first part of this paper will show that the class of strongly convex functions and thus any more generalized classes that include the strongly convex functions are not a suitable setting for Newton's method; hence, also not suitable for global Newton methods. Unfortunately, this is the setting for the asymptotic convergence proofs mentioned above.

Consider the class $C^1(\mu, \lambda)$ having the following definition. Let F be a continuously differentiable map from a separable real Hilbert space H into itself. The inner product in H will be denoted by [,]. Let D(x) denote the Frechet derivative of F at x. By $C^1(\mu, \lambda)$ we denote the set of all maps F for which $\mu\|h\| \leq \|D(x)h\| \leq \lambda\|h\|$ for all h, x $\in H$, with $\mu > 0$. Let $Q = \frac{\lambda}{\mu}$. Assume that the linear operator D(x) has an inverse. We shall show that no global Newton method (or any other algorithm) can do better than linear convergence at a certain determined rate over every member of the above class. Any algorithm that can achieve this rate is called an optimal algorithm. For the special case of $C(\mu, \lambda)$ a simple algorithm due to Nesterov (1983)is optimal to within a multiplicative constant. The convergence rate is linear. Nesterov's algorithm does not require inversions; –it is similar to the gradient method. Any application of Newton's method requires the computation of an inverse operator or the solving a system of linear equations. If the dimension of H is small we usually are willing to pay the price of solving equations to gain the possibility of quadratic convergence. The convergence estimates for the global Newton method in the space $C^1(\mu, \lambda, L)$ that is a subset of $C^1(\mu, \lambda)$ with D(x) satisfying a uniform Lipschitz constant L show arbitrarily slow convergence for sufficiently large values of L. For the special case when D(x) is everywhere self-adjoint we exhibit a gradient algorithm whose efficiency is insensitive to L . For this case the estimate for the gradient method is superior to that of the global Newton method when L is sufficiently large. However the gradient method is sensitive to Q, while the global Newton method is not. Thus for fixed L and large enough values of Q the situation is reversed.

It is a pleasure to thank Brad Bell for discussions and helpful criticisms.

REMARK 1. a). If F is $\in C^1(\mu, \lambda)$ then any stationary point of $\|F(x)\|^2$ is a root of F.

PROOF Let f(x) = [F(x),F(x)]. The differential $f'(x,h)$ = 2[F(x),D(x)h], where D(x)h = $F'(x,h)$. Let h = $D^{-1}(x)F(x)$. If x is stationary then $f'(x,h) = 0 =$ 2[F(x), F(x)]. Whence F(x) = 0.

b).**EXISTENCE.** In view of 1. above, if in addition f has compact level sets then F has roots.

Let $C(\mu, \lambda)$ denote the set of twice differentiable convex functions with

$$\mu\|h\|^2 \leq f''(x,h,h) \leq \lambda\|h\|^2$$

for all x and h $\in$ H, and some positive $\mu \leq \lambda$. The class $C(\mu, \lambda)$ is called a set of "strongly convex" functions. The number $Q = \lambda/\mu$ is called the *condition number*.

ALGORITHMS By an algorithm A(g) where g $\in C(\mu, \lambda)$ we mean a recurrence relation that calculates x_{k+1} using some of the values of g, g' *and* g'' *at* x_s, s=0,1,2,...,k, with x_0 arbitrarily given. A(g) is a special case of a "local method" defined by Nemerovsky and Yuden, 1981. By an algorithm B(F) defined on $C^1(\mu, \lambda)$, we mean a recurrence relation that calculates x_{k+1} using some of the values of F and F' at x_s, $0 \leq s \leq k$, with x_0 arbitrarily given. B(F) is also an instance of a local method. We shall assume that all global Newton methods are B(F) algorithms.

Let $C_s(\mu, \lambda)$ denote a subset of $C^1(\mu, \lambda)$ for which D(x) is self-adjoint with spectral bounds μ *and* λ for all x $\in$ H. For F $\in C_s(\mu, \lambda)$ we can associate a "potential" function f (Vainberg 1955) such that $\nabla f(x)$ = F(x) for all x in H. (Actually, an equivalence class of functions, differing from each other by a constant). Also for every $f \in C(\mu, \lambda)$ there corresponds a $F \in C_s(\mu, \lambda)$. The function f is weakly lower semi-continuous and the level sets of f are weakly sequentially compact. This, and the strong convexity of f implies that there exists a unique minimizer for f, say z.

When any formula below is followed by the word "steps" we mean that the formula is to be rounded up to the nearest integer. We rely on the following claim.

THEOREM 1.)NEMEROVSKY-YUDEN(1979) Given a positive $\epsilon < 1$, a fixed but arbitrary point $x_0 \in H$ and an algorithm A(f), there exists a function f $\in C(\mu, \lambda)$ such that if x_k generated by A(f) reduces $(f(x_k) - f(z))$ to less than $\epsilon(f(x_0) - f(z))$ then k exceeds the number:

$$c\big[\min(\mathrm{n}, \sqrt{\mathrm{Q}})/(\ln\min(\mathrm{n}, \sqrt{\mathrm{Q}}))\big]\ln\frac{1}{\epsilon} = \mathrm{R}\ln\frac{1}{\epsilon}\ \text{steps}$$

Here c is a positive constant, Q is ≥ 2, and z = argmin f.

REMARK 2. For n and Q sufficiently large and ϵ sufficiently small the above bound may be increased to:

$$c\sqrt{Q}\ \ln\frac{1}{\epsilon}$$

Given a positive $\epsilon < 1$ and any function f $\in C(\mu, \lambda)$ there is an algorithm A that yields $f(x_k) - f(z) \leq (f(x_0) - f(z))\epsilon$ whenever k exceeds $4\sqrt{Q}\,(\ln 2)^{-1}\ \ln\epsilon^{-1} = R'\ \ln\epsilon^{-1}$. This algorithm is due to Nesterov(1983). It is essentially optimal, and can only be improved by a decrease in the constant factor $4/\ln 2$. Stated otherwise, the algorithm A applied to any function f $\in C(\mu, \lambda)$ generates a sequence x_k that satisfies $(f(x_k) - f(z))/(f(x_0) - f(z)) \leq ((e^{-(1/R')})^k$, for $1 \leq k \leq \infty$.

The algorithm A (that will be called GRAD1 below) may also be taken to be the gradient method with step length $1/\lambda$. This algorithm requires no information about the values of the function f, while Nesterov's does. Observe that the linearly converging sequence $(e^{-(1/R)})^k$, $k = 1, 2, 3, ...$ is for each k a lower bound for the relative decrease of some function f in $C(\mu, \lambda)$ at x_k , while the sequence $(e^{-(1/R')})^k$ is an upper bound for the relative decrease for any function in $C(\mu, \lambda)$. For the gradient method above the sequence is $(e^{-(1/Q)})^k$.

This prompts us to call the class $C(\mu, \lambda)$ "esslinearly convergent", that is every function in the class can be made to converge no slower than linearly, but sup $\{(f(x_k) - f(z))/(f(x_0) - f(z)) : f \in C(\mu, \lambda)\}$, k=1,2,3,....cannot converge faster than linearly. For brevity we shall refer to this latter property as "sublinearly convergent". We now observe that the class $C_s(\mu, \lambda)$ is also esslinearly convergent.

LEMMA 1 Given F $\in C_s(\mu, \lambda)$, let f denote any potential function for F. Let z = argmin f. The following inequalities obtain:

$$\begin{aligned}(2Q)^{-1}[(f(x) - f(z))/(f(x_0) - f(z))]^{1/2} &\leq \|F(x)\|/\|F(x_0)\| \\ &\leq 2Q[(f(x) - f(z))/(f(x_0) - f(z))]^{1/2} \qquad (A)\end{aligned}$$

Moreover, if

$$f(x) - f(z) \leq \epsilon^2\,(f(x_0) - f(z))/4Q^3 \ \ then\ \ \|F(x)\| \leq \epsilon\,\|F(x_0)\| \qquad (B)$$

PROOF By the strong convexity of f and Taylor's theorem we get:

$$\frac{\mu}{2}\|x-z\|^2 \leq f(x)-f(z) \leq \frac{\lambda}{2}\|x-z\|^2 \qquad (a)$$

By the generalized mean value theorem and the convexity of f we get:

$$\|F(x)\| \leq \lambda\|x-z\| \qquad (b)$$

and

$$f(x)-f(z) \leq \|F(x)\|\,\|x-z\| \qquad (c)$$

By (a) and (b) we have:

$$\frac{\|F(x)\|}{\|F(x_0)\|} \leq \lambda\left[\frac{2(f(x)-f(z))}{\mu\|F(x_0)\|^2}\right]^{1/2} \qquad (d)$$

By (a) and (c) we get

$$\|F(x)\| \geq \frac{\mu}{2}\|x-z\| \qquad (e)$$

and

$$\|F(x)\| \geq \sqrt{\frac{\mu}{2}}\,(f(x)-f(z))^{1/2} \qquad (f)$$

To prove (B) we find that using the hypotheses of (B) together with (d) that

$$\frac{\|F(x)\|}{\|F(x_0)\|} \leq \lambda\left[\frac{2\epsilon^2(f(x_0)-f(z))}{4\mu Q^3\|F(x_0)\|^2}\right]^{1/2}$$

Using (e) we find that the right hand side is less than or equal to

$$\lambda\left[\frac{2\epsilon^2(f(x_0)-f(z))}{4\mu Q^3\mu^2\|x_0-z\|^2/4}\right]^{1/2}$$

Now using (a) the above expression is less than or equal to ϵ.

We now turn to the proof of (A). Using (f) and (d) we find that

$$\sqrt{\frac{2}{\mu}}\,\frac{\|F(x)\|}{\|F(x_0)\|} \geq \frac{(f(x)-f(z))^{1/2}}{\|F(x_0)\|} \geq \left(\frac{f(x)-f(z)}{f(x_0)-f(z)}\right)^{1/2}\frac{\sqrt{\mu}}{\sqrt{2}\,\lambda}$$

. This proves the left side of (A). The right hand inequality is proved similarly, using (d) and (f).

LEMMA 2 The class $C_s(\ \mu,\lambda)$ is esslinear.

PROOF Let f be a potential function corresponding to F. Every algorithm B(F) is now also an algorithm A(f). Every function $F \in C_s(\mu,\lambda)$ is the gradient of some $f \in C(\mu,\lambda)$. Hence for some F , $\|F(x_k)\| / \|F(x_0)\|$ converges more slowly than $(e^{-(1/R)})^{k/2}/2Q$. Now take for the algorithm B the gradient algorithm mentioned above. Again by LEMMA 1 every function F will converge under B with at least a linear rate.

Since $C_s(\mu,\lambda)$ is a subset of $C^1(\mu,\lambda)$, then for some F $\in C^1(\mu,\lambda)$, $\|F(x_k)\| / \|F(x_0)\|$ converges more slowly than $(e^{-(1/R)})^{k/2}/2Q$. Now for B(F) we take the algorithm GRAD2 below. This algorithm converges linearly. Whence we have

THEOREM 2. The class $C^1(\mu,\lambda)$ is sublinearly convergent.

We now restrict the class $C^1(\mu,\lambda)$ to enlarge the possibility of faster convergence. Let $C^1(\mu,\lambda,L)$ denote a map F$\in C^1(\mu,\lambda)$ for which $\|D(x)-D(y)\| \leq L\,\|x-y\|$, for all x,y $\in$ H. The following well- known theorem is adjusted for our present setting.

THEOREM 3 KANTOROVICH(1948) Take $x_0 \in H$.. Let $\beta(x_0) = \|(D^{-1}(x_0))\|$ and $\eta(x_0) = \|(D^{-1}(x_0))F(x_0)\|$. Assume that $\|D(x)-D(y)\| \leq \Lambda(x_0)\|x-y\|$ for all pairs x,y in the ball $B(x_0) = \{x \in H : \|x-x_0\| \leq 2\eta(x_0)\}$. If $\eta(x_0)\beta(x_0)\Lambda(x_0) = h(x_0) \leq 1/2$, then F has a root z such that z is in the ball $B(x_0)$, the Newtonian iterates x_j defined by $x_{j+1} = x_j - D^{-1}(x_j)F(x_j)$ lie in $B(x_0)$, and $\|x_j - z\| \leq 2^{1-j}(2h(x_0))^{2^j-1}\eta(x_0)$.

A convenient terminology similar to Smale's is that under the above circumstances x_0 is an "approximate root".

In what follows we shall take $h(x_0) = 1/4$.

REMARK 3. The condition for an approximate root, $\eta(x_0)\beta(x_0)\Lambda(x_0) \leq 1/4$ has the equivalent condition for an approximate root as:

$$\|F(x_0)\| \leq \alpha(x_0) = 1/[4\beta(x_0)\Lambda(x_0)\|D^{-1}(x_0)F(x_0)/\|F(x_0)\|\,\|]$$

.

REMARK 4. We have for all x $\in$ H global estimates for $\eta(x)$, $\beta(x)$, *and* $\Lambda(x)$. Namely: $\beta(x) \leq 1/\mu$, $\eta(x) \leq \|F(x)\|/\mu$ *and* $\Lambda(x) \leq L$. From these estimates we get:

$$\alpha(x) \geq \mu^2/4L = \alpha$$

$$If \quad \|F(x)\| \leq \alpha \ then\ x\ is\ an\ approximate\ root,\ and$$

$$\eta(x) \leq \mu/4L.$$

In many problems α is so small that the desired accuracy tolerance is achieved before an approximate root is achieved. Thus the efficiency of a global Newton method in reaching an approximate root is a crucial question. We now turn to our version of Smale's algorithm which we shall denote by "SGN". In what follows $\alpha(x_i)$ will be denoted by α_i.

REMARK 5. In what follows the constants $\mu, \lambda,$ and L need not be finite over the entire space H, but rather on the set $S = \{x \in H : \|F(x)\| \leq \|F(x_0)\|\}$.

LEMMA 3. Assume $F \in C^1(\mu, \lambda)$ and x_0 is arbitrarily given in H. If $\|F(x_0)\| \leq \alpha_0$ then x_0 is an approximate root. If not we define a sequence $x_0,\ t_1,\ x_1,\ t_2, \ldots..$ inductively as follows. Given x_i set

$$t_{i+1} = \frac{\|F(x_i)\| - \alpha_i}{\|F(x_i)\|} \qquad (a)$$

Choose x_{i+1} to satisfy

$$\|F(x_{i+1}) - t_{i+1}F(x_i)\| \leq \alpha_i/2 \qquad (b)$$

Then

$$\|F(x_{i+1})\| \leq \|F(x_0)\| - (i+1)\alpha/2$$

where α is defined as in Remark 4.

PROOF. We show first that x_{i+1} can be chosen to satisfy (b). Let $G_i(x) = F(x) - t_{i+1}F(x_i)$. Since $G_i(x_i) = \alpha_i$, x_i is an approximate root for G_i, because $F' = G'$. A few Newton steps (we count them below) suffices to obtain x_{i+1} such that $\|G_i(x_{i+1})\| \leq \alpha_i/2$. Thus (b) can be satisfied. Using the triangle inequality on (a) together with (b) we get that $\|F(x_{i+1})\| \leq \|F(x_i)\| - \alpha_i/2$. Whence $\|F(x_{i+1})\| \leq \|F(x_0)\| - \frac{1}{2}\sum_{j=0}^{i} \alpha_j \leq F(x_0) - (i+1)\alpha/2$. Now choose i so that $\|F(x_{i+1})\| \leq \alpha$

CLAIM 1 Let N be the least integer exceeding $2(\|F(x_0)\| - \alpha)/\alpha$. Then for some $i < N$, x_i is an approximate root of F.

We now estimate the number of Newton steps to move from x_i to x_{i+1}.

LEMMA 4 Let $\{y_{ij}\}$ be a sequence of Newtonian iterates starting at $y_{i0} = x_i$. Let $G_i(z_i) = 0$. Then we can choose $x_{i+1} = y_{iK}$ where K is the least integer $\geq 1.443 \ln(1.443 \ln 8Q)$.

PROOF We have seen that x_i is an approximate zero for G_i hence $\|y_{ij} - z_i\| \leq \frac{\mu}{L}(\frac{1}{2})^{2^j}$. Then

$$\|G_i(y_{ij}) - G_i(z_i)\| \leq \lambda\|(y_{ij} - z_i\| \leq \lambda\mu L^{-1}\left(\frac{1}{2}\right)^{2^j}.$$

Now choose K so that $\lambda\mu L^{-1}(\frac{1}{2})^{2^K} \leq \alpha/2$, that is $(\frac{1}{2})^{2^K} \leq 1/(8Q)$.

REMARK 6. The above algorithm can be optimized by changing the right hand side of inequality (b) in the recursion above to α/q with $q > 1$. The formula for N becomes $(\|F(x_0)\| - \alpha)q/\alpha)$ and K becomes $1.433 \ln(1.433 \ln 4qQ)$. Now choose q to minimize NK.

LEMMA 5 Take F $\in C^1(\mu, \lambda, L)$ Assume that D(x) is self-adjoint for all x $\in$ H. The gradient method previously mentioned below REMARK 2, called Algorithm A, that we shall now call "GRAD1" will, starting at x_0, generate an approximate root in K steps, where K is the smallest integer $\geq Q \ln[\|F(x_0)\| 4QL/\mu^2]$

PROOF The mapping G defined by $G(y) = y - F(y)/\lambda$ has a fixed point z satisfying F(z) = 0. It is a contractor satisfying a Lipschitz condition $q = 1 - 1/Q$. Goldstein(1967, pps 15 and 24). Set $G(x_n) = x_{n+1} = x_n - F(x_n)/\lambda$. Then $\|F(x_n) - F(z)\| \leq \lambda \|x_n - z\| \leq \lambda q^n \|x_0 - x_1\|/(1-q) = \|F(x_0)\| q^n/(1-q)$. Now choose n so that

$$\|F(x_0)\| q^n \leq \alpha(1-q)$$

Using the inequality $-\ln(1 - \mu/\lambda) \geq \mu/\lambda$, one obtains the lemma.

We now consider a gradient method for the non-symmetric case. We call this gradient method GRAD2.

ALGORITHM GRAD2 Take F $\in C^1(\mu, \lambda, L)$, $x_0 \in S$ and set f(x) = $\|F(x)\|^2$. Then $\nabla f(x)$ satisfies $\|\nabla f(x) - \nabla f(y)\| \leq M\|x - y\|$ for all x and y in S, with $M = 2(\lambda^2 + \|F(x_0)\| L)$. Set $\phi(x) = f(x)\nabla f(x)/\|\nabla f(x)\|^2$. Given arbitrary x_0 in H set $x_{k+1} = x_k - \gamma_0 \phi(x_k)$ with $\gamma_0 = \mu^2/2M$. If k exceeds

$$2\left(\frac{\|F(x_0)\|\, L}{\mu^2} + Q^2\right) \ln\left(\frac{\|F(x_0)\|\, 4L}{\mu^2}\right)$$

then x_k is an approximate root.

PROOF Adding and subtracting $(F'(y))^* F(x)$ we find that $\|\nabla f(x) - \nabla f(y)\| \leq 2(\lambda^2 + \|F(x)\|\, L)\, \|x - y\|$, and $f(x) - f(x - \gamma\phi(x)) = \gamma[\nabla f(x), \phi(x)] + \gamma[\nabla f(x) - \nabla f(\xi), \phi(x)]$. Here $\|\xi - x\| \leq \gamma M \|\phi(x)\|$. Then $f(x - \gamma\phi(x)) \leq f(x) - \gamma f(x) + \gamma^2 M \|\phi(x)\|^2$, $\nabla f(x) = 2(F'(x))^* F(x)$, and $\|\nabla f(x)\| \geq 2\|F(x)\|\mu$. Then $f(x_{k+1}) \leq f(x_k)[1 - \gamma_0 + \gamma_0^2 M/4\mu^2] = f(x_k)(1 - \mu^2/M)$. Taking square roots we get that $\|F(x_{k+1})\| \leq \|F(x_k)\|(1 - \mu^2/2M)$. Finally, choose k so that $\|F(x_0)\|(1 - \mu^2/2M)^k \leq \mu^2/4L$,

Comparing the algorithms. Let

$$S = \frac{\|F(x_0)\| L}{\mu^2}$$

By claim 1 and lemma 4 the total number of steps of SGN is:

$$SGN \ : \ 11.544(S - .25)\ln(1.443\ln 8Q).$$

$$GRAD2 \ : \ 2(S + Q^2)\ln 4S$$

$$GRAD1 \ : \ Q\ln(4SQ)$$

Notice that unlike GRAD1 and GRAD2, SGN is insensitive to the condition number Q! However SGN is sensitive to S. GRAD 1 is sensitive to Q but not to S. GRAD2 is sensitive to both of these factors. In the symmetric case for fixed Q, GRAD1 is quicker than SGN when $\|F(x_0)\|$ grows sufficiently large or if L/μ^2 gets sufficiently large. On the other hand for fixed S, SGN is quicker as Q grows sufficiently large. In the non-symmetric case SGN is superior to GRAD2 with respect to the number of steps. When the cost per step is included, the gradient methods become cheaper in the n-dimensional case when n is sufficiently large. For each Newton step an nxn system of linear equations is solved, costing $O(n^3)$ multiplications. While the corresponding GRAD2 step involves a matrix multiplication of an nxn and a nx1 matrix, or n^2 multiplications, and GRAD1 requires no matrix operations.

Bibliography

J. Abadie and G. Guerrero.(1984) Methode du GRG,Methode de Newton Globale et application a la Programmation Matematique.RAIRO Recherche Operationelle vol 18,no 4, p319-351.

D. Bertsekas.(1982) Projected Newton Methods for Optimization Problems with Simple Constraints, SIAM J. Control. Opt.,20, 2, pp. 221-246

J.C. Dunn(1980) Newton's method and the Goldstein step length rule for constrained minimization problems. SIAM J Control Opt 18:659-674.

J.C. Dunn.(1986) Projected Newton Methods for Nonlinearly-Constrained Minimization Problems. To appear.

C.B. Garcia and F.J. Gould(1979), Scalar labelings for homotopy paths, Math. Programming,17 pp184-197.

C.B Garcia and F.J. Gould(1980),Relations between several path following algorithms and

global Newton methods,SIAM REVIEW, 3,pp264-273.

A.A. Goldstein.(1965) On Newton's method, Numerische Math. 7,391-393(1965).

A.A. Goldstein(1967) Constructive Real Analysis, Harper and Row, N,Y., N.Y.

G.C. Hughes and J.C. Dunn(1984), Newton-Goldstein Convergence Rates for Convex Constrained Minimization Problems with Singular Solutions,Appl Math Optim 12:203-230

H.T. Kung(1976) The Complexity of Obtaining Starting Points for Solving Operator Equations by Newton's Method. Analytic Computational Complexity. Edited by J.F. Traub. Academic Press, N.Y.

A.S. Nemerovsky and D.B. Yudin(1979) Problem Complexity and method efficiency in optimization.Translated(1983) by E.R. Dawson, Wiley, 388pp

Yu E. Nestorovich A Method of Solving a Convex Programming Problem with Convergence Rate $O(1/k^2)$. Doklady Nauk, Vol 269(1983) no. 3; Soviet Math. Dok.,vol 27, 1983.

J.M. Ortega and W.C. Rheinboldt(1970) Iterative Solution of Nonlinear Equations in Several Variables. Academic Press, N.Y. 571pp

B.N. Pshenichnyi(1970) Newton's method for the solution of equalities and inequalities. Math. notes of the Academy of Sciences,USSR 8 pp827-830.

E. Polak(1973) A historical survey of computational methods in optimal control. Twentieth Anniversary of SIAM, Special Lectures. SIAM REVIEW15 553-584

S.M Robinson(1972) Extension of Newton's method to nonlinear functions with values in a cone. Num. Math. 19 341-347.

S.Smale(1976) A convergent process of price adjustment and global Newton Methods, J. Math Econom.3,107-120

S. Smale(1985) On the efficiency of algorithms of analysis, Bull Amer. Math. Soc. (N.S.) 134,1-36.

S. Smale(1986-2) Algorithms for solving equations. Invited Address, International Congress of Mathematicians, Vol 1 pp 172,195 AMS 1986.

M.M. Vainberg(1955) Variational Methods for the Study of Nonlinear Operators. English

version by Holden Day 1964.

D.B. Yudin and A.S. Nemerovsky(1977) Computational complexity of strictly convex programming, Ekonomika i matem. metody XIII, No3 550-569 (Russian)

INITIAL DISTRIBUTION LIST

DIRECTOR (2)
DEFENSE TECH. INFORMATION
CENTER, CAMERON STATION
ALEXANDRIA, VA 22314

DIRECTOR OF RESEARCH ADM.
CODE 012
NAVAL POSTGRADUATE SCHOOL
MONTEREY, CA 93943

OFFICE OF NAVAL RESEARCH
CODE 422AT
ARLINGTON, VA 22217

PROF. ALLEN GOLDSTEIN (10)
CODE 53GO
DEPARTMENT OF MATHEMATICS
NAVAL POSTGRADUATE SCHOOL
MONTEREY, CA 93943

LIBRARY (2)
NAVAL POSTGRADUATE SCHOOL
MONTEREY, CA 93943

DEPT. OF MATHEMATICS
CODE 53
NAVAL POSTGRADUATE SCHOOL
MONTEREY, CA 93943

CENTER FOR NAVAL ANALYSES
4401 FORD AVENUE
ALEXANDRIA, VA 22302-0268